DOCUMENT RESUME

ED 180 191 FL 010 668

AUTHOR Afendras, Evangelos A.
TITLE Can One Measure a Sprachbund? A Calculus of Phonemic Distribution for Language Contact.
INSTITUTION Laval Univ., Quebec (Quebec). International Center for Research on Bilingualism.
PUB DATE 69
NOTE 19p.; Paper presented at meeting of the Societas Linguistica Europaea (Vienna, Austria, 1969)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Bilingualism; Dialects; *Dialect Studies; Ethnology; *Language Patterns; Language Research; Language Typology; *Language Variation; *Linguistic Borrowing; *Mathematical Linguistics; *Models; Morphology (Languages); Multilingualism; Phonemes; Phonology; Sociolinguistics
IDENTIFIERS Balkans; *Language Contact

ABSTRACT

Language contact and the resulting interference has long been diagnosed as one of the primary forces behind language change. In cases of multilingual contact within geographically restricted areas, converging changes of the languages in contact have been uncovered and described. The geographic areas characterized by such linguistic situations came to be known as "Sprachbundes" or Language Convergence Areas. The need for an integrated, interdisciplinary approach to convergence led to a search for the quantification of the linguistic dimension. The classic example of a language area has been the Balkan Peninsula, where language, ethnicity, nationalism, politics, religion, and economy have produced a societal variety and some similarities. In this project, several Balkan languages were analysed and the matrices for two-phoneme combinations were constructed by an exhaustive examination of the available lexica. From these matrices were calculated the indices, completeness, symmetry, reflexivity, and transitivity according to the mathematical formulas derived by Farary and Paper. The results indicate that the method can be considered a very efficient one: it establishes and measures convergence with results congruent with those of a totally different model, and it can be applied to any convergence area. (AMH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED180191

CAN ONE MEASURE A SPRACHBUND? A CALCULUS OF PHONEMIC DISTRIBUTION FOR LANGUAGE CONTACT *

by Evangelos A. Afendras

International Center for Research on Bilingualism

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

I.C.R.B

Alain Pregnier
Acting Director

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL010 668

1969

*The original version of this paper was presented at the *Societas Linguistica Europaea*, 3. Jahrestagung, Vienna, September 6th, 1969, with the title *Quantification of Convergence in Balkan Phonological Distribution*.

Publication B-22

CONTENTS

0. Introduction
1. The Need for a Metric
2. The Model
3. The Balkan Case
4. The Results
5. Conclusions and Suggestions

Appendices
Tables
Bibliography

SUMMARY

In the quest for measures of language convergence (and some aspects of Contact), the mathematical model of phoneme distribution based on two-phoneme sequences proposed by Harary and Paper (1957) was found appropriate.

As a test, some Balkan and non-Balkan languages were analysed and their properties were statistically correlated and plotted.

Revealing, partly unexpected findings suggest that the model is successful and should be applied to other cases. Finally, some suggestions are made for broadening the data base and facilitating its extraction.

0. Introduction

Language contact, including contact between closely related dialects or even styles, and the resulting interference has long been diagnosed as one of the primary forces behind language change. In cases of multilingual contact within geographically restricted areas, Jakobson and others have uncovered and described converging changes of the languages in contact which result in a situation of typological homogeneity replacing genetic heterogeneity. The geographic areas characterized by such linguistic situations came to be known as *Sprachbünde* (Trubetzkoy, 1928) and in the Anglo-American Jargon as *Language Convergence Areas*.

1. The Need for a Metric

Several problems are posed in the investigation of such areas, investigations necessary for the understanding of the change processes resulting from bilingualism. Among them the following:

1.1 establishing the existence and origin of features typical of the area;

1.2 establishing the degree of convergence.

The first is a question of typology, individual language history and typology of change. In other words, one has to detect similarities in the language structures involved, trace their development in these languages and establish possible diffusionary spread from one or more source languages to the others. Then, the researcher must evaluate the extent to which the "adopting" structures were "bent" in order to accept the structural "suggestion".

The second problem mentioned above is again based on linguistic typology and can be most efficiently approached with a measuring method. One can hardly disagree on this point. For those, however, who would question the overall usefulness of such a study, let me introduce a third major problem, namely, understanding the sociolinguistic process of language convergence (1.3). For obvious reasons, the sociolinguistic component represents the heart of a multilateral contact study. But unless one can quantify the linguistic component which is directly affected and certainly the part of the cultural exchange most amenable to measure, the usefulness of the sociolinguistic study is seriously impaired.

The necessity for an integrated, interdisciplinary approach to convergence has lead me to a search for the quantification of the linguistic dimension.

2. The Model

In the 8th International Congress of Linguistics (1957), Haugen and Weinreich contributed a stimulating dialogue on Bilingualism, Contact and Convergence. A model for quantification of phonemic distribution (Harary-Paper, 1957) was included in Weinreich's suggestions for a precise analysis of interference. The original authors illustrated this model with the analysis of two Japanese dialects and since that time its appropriateness has been discussed by linguistic theoreticians, mathematical linguists and typologists (see Bibliography - Haarg-Hanssen, Herdan, Revzin, Buzássyová, etc.).

In the model under discussion, some phonotactic properties deriving from permissible two-phoneme sequences are examined. A matrix is constructed with the phonemes of the language as its two dimensions (= N) and with N^2 cells. An empty cell expresses the absence of that particular ordered phoneme sequence. This can be the result of chance, but could also be the result of structural limitations. The phonemes

are now considered as elements in a closed set which enter (into) a binary relation R; e.g., if we have elements (a,b), the relation R means that the ordered sequence /ab/ is permissible and occurs. The matrix of these ordered sequences for a given language has several mathematical properties.

2.1 The completeness of the field directly expresses the number of existing binary combinations; it is a measure of the degree to which the systems utilize their resources; or, conversely, a partial index of structural limitations. (The gaps in the matrix can be both structural and accidental).

2.2 Reflexivity examines the extent to which gemination is allowed; (gemination and length are not to be confused).

2.3 Symmetry expresses the importance of order for the binary combinations. E.G. if the elements (x,y) occur in the ordered sequence /xy/, does the ordered sequence /yx/ also occur? This measure is, along with completeness, a measure of distributional freedom.

3.4 Transitivity, one more index, introduces a slightly more complicated notion. For each element x, there are two classes of elements related to x as follows:

class 1 includes all the elements with which x combines;

class 2 includes all the elements with which the elements of class 1 combine.

Then, the degree of transitivity reflects the extent to which element x combines with members of class 2.

3. The Balkan Case

The classic example of a language area has been the Balkan Peninsula, where language, ethnicity, nationalism, politics, religion and economy have produced a fascinating societal variety and within it, some even more fascinating similarities.

The idiomatic, morpho-syntactic, and lexical levels have been presented in Sandfeld's model study (*Linguistique Balkanique: Problèmes et Résultats*, Paris, 1932) and in some powerful recent efforts (Birnbaum 1966, 1968 a, b, 1969; Klagstadt 1958; Kazazis 1964, 1967). The phonology, although often used for purposes of illustration by Jakobson and Trubetzkoy, and outlined in an early effort by Bohuslav Havránek (1933), has only recently been the subject of more systematic study (Hamp 1962, Afendras 1968, 1969 a, b). In another study (1969 b), I investigated convergence reflected in the distinctive feature utilization of 51 Balkan vocalic systems. As will be seen later, the results of the distinctive feature analysis support the results of the present study.

In this project, several Balkan languages were analysed and the matrices for two-phoneme combinations were constructed by an exhaustive examination of the available lexica. This is a time consuming process and could not be repeated for all the languages and dialects of the Balkans. From these matrices were calculated the indices listed above, Completeness, Symmetry, Reflexivity and Transitivity, according to the mathematical formulas derived by Harary and Paper. The final step involved cross-language comparison. I correlated the languages analysed for their four indices, K(R), S(R), G(R), T(R), and then submitted the correlation matrices to multidimensional scaling, and clustering analyses (Fig. 4).

4. The Results

For the small number of languages included (Arumanian, Makedonski, two varieties of Standard Greek, Sarakačan Greek, Old Church Slavic, Attic Greek and two dialects of Japanese) the results were rather surprising. In some of the plots the languages were arranged in a meaningful time order. For instance, the one South Slavic language included (Makedonski), was grouped with Modern Greek, Arumanian and, most surprising, Attic Greek, away from Slovak and even farther away from Old Church Slavic. One could tentatively conclude from these results that some South Slavic languages have changed from the type with the phoneme combinability or syllabic structure to which belonged OCS, to a type common in the Balkans and characteristic of Greek since the time of Classical Attic. As might have been expected (although the corpus available for determining the binary matrix was insufficient), Sarakačan differs greatly from the rest of Greek as a result of its loss of unstressed /i/ and /u/. Very basic laws determining the nature of Greek (and even Indo-European) consonant clusters are violated in Sarakačan, but some, e.g. the regressive assimilation for voicing throughout a cluster, are kept. As can be seen from the plotted diagram, the various Balkan languages and Attic Greek phonemic systems are even closer related than the two Japanese dialects are related to each other.

5. Conclusions and Suggestions

The method can be considered to be a very efficient one. It establishes and measures convergence with results congruent with those of a totally different model (distinctive feature utilization in vocalic systems) in which, again, Attic is found to be closer to

more Balkan languages than Old Church Slavic or Common Slavic. The results could not have been easily predicted from a non-mathematical analysis of the linguistic structures. Two things must now be developed: the statistical analysis must be grasped stronger so as to be closer tied to linguistic explanations (this both on the level of the indices and their subsequent statistical correlation); then, as extraction of the binary phoneme sequence matrix is such a long and painstaking process, some automation must be used:

1) for each language a correspondence between graphemes and phonemes should be established;

2) a phonemic transcription of a given lexicon (or a given dictionary) could then be produced automatically;

3) a concordance of phonemes could be produced next, from which could finally be abstracted a data matrix.

The following steps of comparison, correlation, etc., are the same as in the present study. This task is by no means impossible since an algorithm for automatic phonemic transcription of Czech, Russian and German texts (or lexica) has already been devised and used (Kučera-Munroe, 1968).

By applying the method outlined here any convergence area, be it India, the Caucasus, Colombia, Nigeria or Canada, could be analysed and the results lent to sociolinguists and others for an integrated, global study.

Appendices

The formulas for the four Harary - Paper indices are the following:

(1) Completeness: $$K(R) = \frac{nR}{(nP)^2 - 1}$$

nP = number of phonemes in the language
nR = number of actual binary sequences in the matrix
R = the relation of ordered binary sequences.

(2) Symmetry = $$S(R) = \frac{1}{2} \frac{\sum_{x \in P} n(\alpha x \Theta \beta x)}{nR}$$

αx = all the predecessors of x (= phonemes preceding x)
βx = all the successors of x (= phonemes following x)
$\alpha x \Theta \beta x$ = the symmetric difference of αx and βx, i.e. all the phonemes which can either follow or precede x but cannot be both.

(3) Reflexivity = $$G(R) = \frac{\sum_{x \in P} Gx}{nP-1}$$

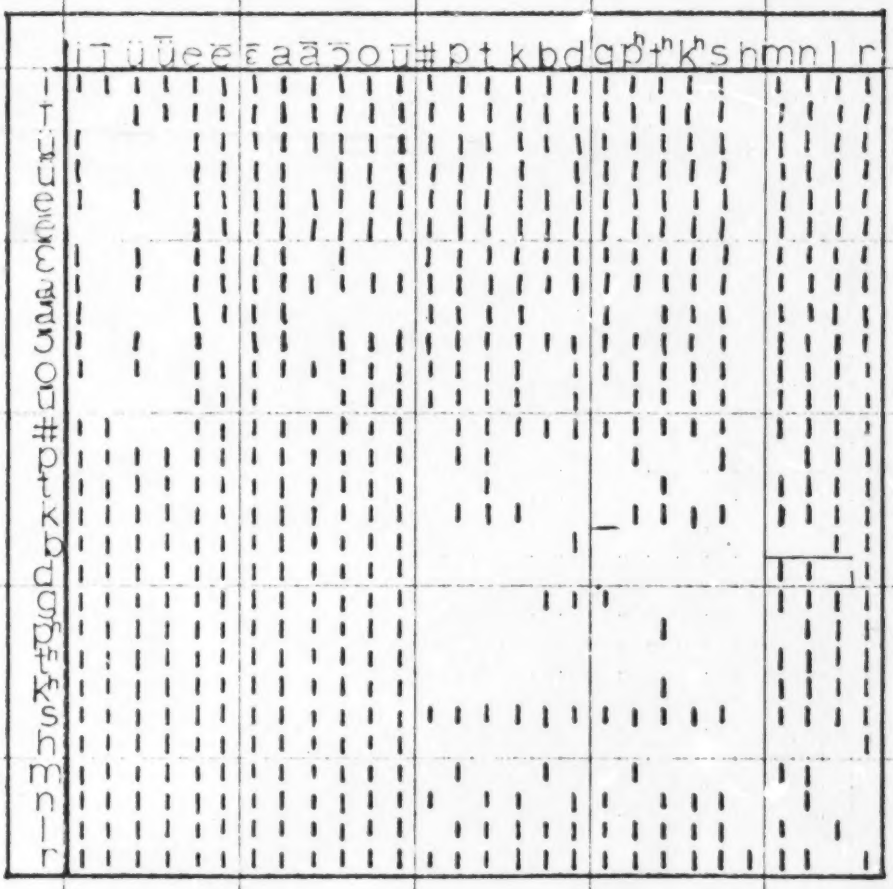
where Gx = 1 if xx occurs,
Gx = 0 if xx does not occur.

If all phonemes were able to combine with themselves, we would have the maximal reflexivity of 1.## (=pause) by definition does not combine with itself.

(4) Transivity = $$T(R) = \frac{\sum_{xy \in R} Txy}{nR}, \quad Txy = \frac{(\beta y \cap \beta x)}{n\beta y}$$

where $\beta y \cap \beta x$ is the intersection of βy and βx.

Binary Phoneme Matrix for Attic Greek (5-4th) Cent. B.C.



ü = ʊ
ɛ̈ = ɜ
ä = ɜ̈
ɔ = ɞ
ɛ = ɳ

REFLEXIVITY (14)= 0.630 IRREFLEXIVITY (15)= 0.370

ORIGINAL DATA MATRIX FOR **ATTIC GREEK** (5TH-4TH CENT. B.C.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
2	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
3	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
4	1	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	
5	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
6	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
7	1	0	1	0	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
8	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
9	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	1	0	1	1	1	0	1	1	1	1
10	1	0	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
11	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
12	0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1
13	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
14	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	1	0	0	1	0	0	1	1	1
15	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	1	0	0	0	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	1	1	1	0	1	1	1	1
17	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1
18	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
19	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	1
20	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1
21	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
22	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
24	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
25	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1	0	0
26	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	1	1	0	0	1	0	0
27	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1	0
28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1

K(R)= 0.752

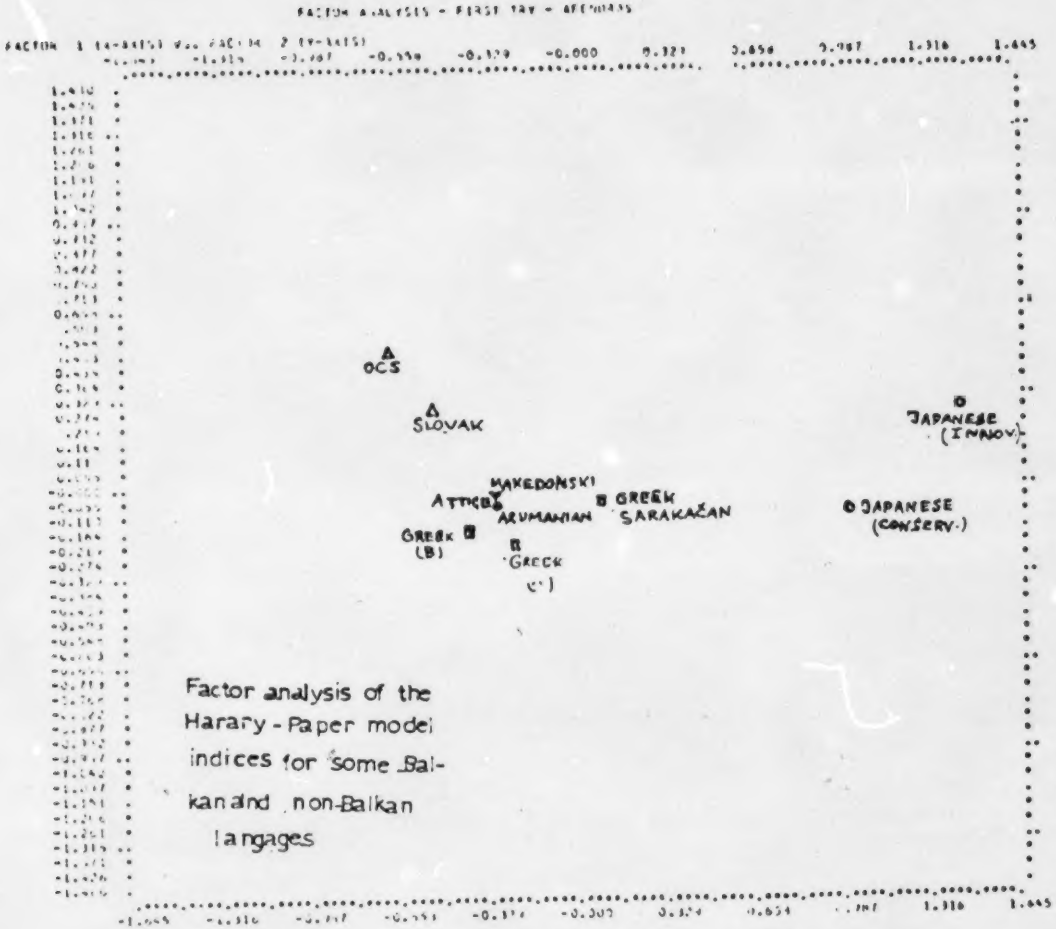
S(R)= 0.837

T(R)= 0.807

11

		K(R)	S(R)	G(R)	T(R)
1.	Arumanian	.461	.735	.079	.647
2.	Attic Greek	.752	.837	.630	.807
3.	Greek A	.525	.764	.154	.740
4.	Greek B	.588	.799	.154	.757
5.	Japanese Conserv.	.390	.730	.430	.760
6.	Japanese Innov.	.380	.750	.530	.750
7.	Makedonski	.574	.823	.226	.741
8.	Old Church Slav.	.478	.684	.118	.459
9.	Sarakačan	.422	.695	.229	.651
10.	Slovak	.513	.750	.130	.565

Completeness, Symmetry, Reflexivity and Transivity measures for ten languages submitted to a Factor Analysis.



Factor analysis of the Harary-Paper model indices for some Balkan and non-Balkan langages

BIBLIOGRAPHY

For a more complete bibliography on Balkan linguistic convergence, see Afendras (1968, 1969b).

1. Afendras, Evangelos A. 1968. The Balkans as a Linguistic Area: A Study in Phonological Convergence. Unpublished Ph. D. Dissertation. The Johns Hopkins University, Baltimore.

2. ———————— 1969a. Stability of a Bilingual Situation and Arumanian Bilingualism. Paper read at the annual meeting of the Canadian Linguistic Association, Toronto, June 1969.

3. ———————— 1969b. Mathematical Models for Balkan Phonological Convergence. Paper presented at the *Third International Congress on Computational Linguistics*, Sänga-Säby, Sweden, September 1-4, 1969. (Preprint No. 56).

4. ———————— 1969c. On Bilingualism and Language Structure: a Review, a Case, a Synthesis, *Culture*, Vol. 30.4 , December 1969.

5. Birnbaum, Henrik 1966 - On Typology, Affinity, and Balkan Linguistics. *Zb. za filologiju i linguistiku*. IX, 17-30. Novi Sad.

6. ———————— 1968 Slavjanskije jazyki na Balkanax i ponjatije tak nazyvajemyx jazykovyx sojuzov. *Glossa*, Vol. 2-1, 70-92.

7. ———————— 1969. Deep Structure and Typological Linguistics. Unpublished manuscript.

8. Bright, William (ed.) 1966. *Sociolinguistics*. Mouton, The Hague.

9. Buzássyová, K. 1966. Calculus of Distribution of the Phonological System. Prague *Studies in Mathematical Linguistics*. I, p. 51-64.

10. Caragiu-Marioţeanu 1964. Phonologie aroumaine. *Revue Roumaine de Linguistique*, IX, 534-547.

11. ———————— 1968. *Fono-Morfologie Aromână: Studiu de Dialectologie Structurală*, Bucureşti.

12. Civ'jan, T.V. 1966. *Imja suščestvitel'noje v balkanskix jazykax (:k strukturno-tipologičeskij xarakteristike balkanskogo jazykovogo sojuza)*. Moscow, Akademija Nauk.

13. Emeneau, M.B. 1956. India as a Linguistic Area. *Lg* 32: 3-16.

14. ———————— 1962. Bilingualism and Structural Borrowing. *Proceedings of the Americal Philosophical Society*. 106.5:430-442.

15. Gołąb, Z. 1959. Some Arumanian-Macedonian Isogrammatisms and the Social Background of their Development. *Word* 15.415-435.

16. Gumperz, John E. 1966. On the Ethnology of Linguistic Change. In Bright 1966: 27-49 (incl. discussion).

17. Harary, F. and H. Paper. 1957. Toward a General Calculus of Phonemic Distribution. *Lg*. 33. 143-169.

18. Haugen, Einar I. 1957. Language Contact. *Proceedings of the VIII International Congress of Linguists*, 1957, 771-785.

19. Herdan, G. 1960. *Type-Token Mathematics. A Textbook of Mathematical Linguistics.* (See Chapter 7.5, which reviews Harary-Paper's Model.)

20. Hymes, Dell H. 1967. Models for the Interaction of Language and Social Setting. *Journal of Social Issues*, Vol. 23. 8-28.

21. Ivić, P. 1968. Liens phonologiques entre les langues balkaniques. In: *Actes du premier Congrès international des études balkaniques et sud-est européennes.* 133-141. Bulgarian Academy of Science. Sofia.

22. ——— 1967. Sur les affinités phonologiques des langues et des dialectes sur les Côtes septentrionales de la Méditerranée. In: *Bolletino del Atlante Linguistica Mediterraneo.* Vol. 8-9, pp. 15-25. Firenze.

23. Jakobson, R. 1931. K xarakteristike jevrazijskogo jazykovogo sojuza. Reprinted in: *Selected Writings I,* 117-136.

24. ——— 1931. Über die phonologischen Sprachbünde, in *S.W. I,* 137-143.

25. ——— 1938. Sur la théorie des affinités phonologiques entre les langues. *S.W. I.* 234-246.

26. ——— 1961. Results of the Congress. *Proceedings of the 9th International Congress of Linguistics,* Mouton, 1962, 1135-42, especially p. 1139.

27. Kazazis, Kostas. 1964. Some Balkan Constructions Corresponding to W. European Infinitives. Unpublished Ph. D. Thesis, Indiana University, Bloomington.

28. ———————— 1967. On a Generative Grammar of the Balkan Languages. *Foundations of Language*. 3. 117-123. Dordrecht.

29. Mackey, W.F. 1965. Bilingual Interference: its Analysis and Measurement, *The Journal of Communication*, 15.4:239-249.

30. Pavlović, Milivoj 1958. Les processus balkanistiques et la perspective du parler serbe de Prizren. *Orbis* 7:46-53.

31. Récatas, Basile 1934. *L'état actuel du bilinguisme chez les Macédo-roumains du Pinde et le rôle de la femme dans le langage.* Paris, Droz.

32. Revzin, I.I. 1962. *Modeli Jazyka.* Moscow.

33. Sandfeld, K. 1932. *Linguistique balkanique: Problèmes et résultats.* Paris.

34. Širokov, O.S. 1964. Voprosy balkanskogo vtoričnogo jazykovogo rodstva (primenenije statistiki v diaxroničeskoj fonologii), in *Problemy sravnitel'noj grammatiki indojevropejskix jazykov. Naučnaja sessija. Tezisy dokladov.* Moscow. 68-69.

35. Spang-Hanssen, H. 1962. Mathematical Linguistics - Trend in Name or in Fact? *FJICL.* The Hague, Mouton. 61-71 (incl. discussion).

36. Trubetzkoy, N.S. 1928. Proposal 16 in First International Congress of Linguists, *Actes* p. 17 f.

37. Weinreich, U. 1953. *Languages in Contact.* New York.

38. ———————— 1957. Research Frontiers in Bilingualism Studies. *Proceedings of the 8th International Congress of Linguistics*, 786-810. (Including discussion.)

39. ———————— 1958. On the compatibility of genetic relationship and convergent development. *Word* 14. 374-379.